

11. <https://github.com/topics/natural-language-to-sql>
 12. <https://github.com/FerreroJeremy/ln2sql>
 13. <https://github.com/rupinder1133/ln2sqlmodule>
 14. <https://github.com/maverickjoy/ln2sql>
 15. https://github.com/joanna-bb/Postgres_LLM/blob/main/Postgres_LLM_Llama_ex1.py
 16. <https://github.com/bhattbhavesh91/google-gemma-finetuning-n2sql/blob/main/n2sql-google-gemma-finetuning-notebook.ipynb>
 17. <https://github.com/bhattbhavesh91/n2sql-google-gemini>
 18. <https://forum.octopus.energy/t/maximising-oversized-array-or-arguably-undersized-inverter/9750/14>
 19. <https://www.marktechpost.com/2024/05/02/this-ai-paper-introduces-llama-3-8b-instruct-80k-qlora-new-horizons-in-ai-contextual-understanding/?amp>
 20. <https://huggingface.co/namespace-Pt/Llama-3-8B-Instruct-80K-QLoRA>
 21. https://huggingface.co/namespace-Pt/Llama-3-8B-Instruct-80K-QLoRA-Merged-GGUF/blob/main/Llama-3-8B-Instruct-80K-QLoRA-Merged-Q4_K_M.gguf
 22. <https://www.unite.ai/decoder-based-large-language-models-a-complete-guide/>
 23. https://www.google.com/search?q=llm+lora+tutorial&oq=llm+lora+&gs_lcrp=EgZjaHJvbWUqBwgFEAAyGAYyBggAEEUYOTIHCAEQABiABDIHCAIQABiABDIHCAMQABiABDIHCAQQABiABDIHCAUQABiABDIHCAYQABiABDIICAcQABgWGB4yCAgIEAAYFhgeMggICRAAGBYHjIICAoQABgWGB4yCAgLEAAYFhgeMggIDBAAGBYHjIICA0QABgWGB4yCAgOEAAAYFhge0gEJMTg2NzVqMGo3qAIUsAIB&client=ms-android-oneplus-rvo3&sourceid=chrome-mobile&ie=UTF-8#ip=1
 24. <https://zohaib.me/a-beginners-guide-to-fine-tuning-llm-using-lora/amp/>
 25. <https://xiaosean5408.medium.com/fine-tuning-llms-made-easy-with-lora-and-generative-ai-stable-diffusion-lora-39ff27480fda>
 26. <https://medium.com/data-science-in-your-pocket/lora-for-fine-tuning-llms-explained-with-codes-and-example-62a7ac5a3578>
 27. <https://www.datacamp.com/tutorial/mastering-low-rank-adaptation-lora-enhancing-large-language-models-for-efficient-adaptation>
 28. <https://www.databricks.com/blog/efficient-fine-tuning-lora-guide-llms>
 29. <https://magazine.sebastianraschka.com/p/practical-tips-for-finetuning-llms>
 30. <https://learnopencv.com/sd-xl-inpainting/>
 31. <https://civitai.com/models/176555?modelVersionId=214296>
 32. <https://civitai.com/tag/text>
 33. <https://venturebeat.com/ai/metasp-new-multi-token-prediction-makes-ai-models-up-to-3x-faster/>
 34. <https://venturebeat.com/>
 35. https://www.google.com/search?q=how+to+embed+factual+knowledge+into+llm&oq=how+to+embed+factual+knowledge+into+llm&gs_lcrp=EgZjaHJvbWUyBggAEEUYOTIHCAEQIRigAdIBCTI3NzcxajBqN6gCFLACAQ&client=ms-android-oneplus-rvo3&sourceid=chrome-mobile&ie=UTF-8
 36. <https://www.marktechpost.com/2024/05/06/nvidia-publishes-a-competitive-llama3-70b-quality-assurance-qa-retrieval-augmented-generation-rag-fine-tune-model/?amp>
 37. <https://huggingface.co/nvidia/Llama3-ChatQA-1.5-8B/discussions/5>
 38. <https://huggingface.co/QuantFactory/NVIDIA-Llama3-ChatQA-1.5-8B-GGUF>
 39. <https://alphaarchitect.com/2024/05/forecast-equity-risk-premium/>
 40. <https://github.com/zylon-ai/private-gpt/pull/1825>
 41. <https://huggingface.co/blog/cost-efficient-rag-applications-with-intel>
 42. <https://huggingface.co/Qwen/CodeQwen1.5-7B-Chat>
 43. <https://huggingface.co/deepseek-ai/deepseek-coder-6.7b-instruct>
 44. <https://future.mozilla.org/news/llamafiles-for-embeddings-in-local-rag-applications/>
 45. https://www.google.com/search?client=ms-android-oneplus-rvo3&sca_esv=1a5929d447859cf0&sxsrf=ADLYWII7aAUDwsLy9GDbpNCso2Djb-OMnA:1715908662447&q=Pytorch+parallel+inference+on+single+GPU&sa=X&ved=2ahUKEwiu6qrywZOGAxXrV0EAHVj2BZIQ1Qj6BAgeEAE&biw=360&bih=663&dpr=3
1. <https://github.com/CopilotKit/CopilotKit>
 2. <https://github.com/OS-Copilot/OS-Copilot>
 3. <https://github.com/ex3ndr/llama-coder>
 4. <https://github.com/Kuingsmile/word-GPT-Plus?tab=readme-ov-file>
 5. <https://github.com/Kuingsmile/word-GPT-Plus?tab=readme-ov-file>

<https://github.com/srikanth235/privy> 6. <https://github.com/rubberduck-ai/rubberduck-vscode> 7.
<https://github.com/Bin-Huang/chatbox?tab=readme-ov-file>

14. <https://www.marktechpost.com/2024/05/19/meet-verba-1-0-run-state-of-the-art-rag-locally-with-ollama-integration-and-open-source-models/?amp> 15.
- <https://github.com/weaviate/Verba?tab=readme-ov-file> 16.
- <https://amp.theguardian.com/technology/article/2024/may/20/microsoft-chatbot-assistant-pc> 17.
- <https://amp.theguardian.com/technology/article/2024/may/20/world-is-ill-prepared-for-breakthroughs-in-ai-say-experts> 19.
- <https://venturebeat.com/ai/the-future-of-financial-analysis-how-gpt-4-is-disrupting-the-industry-according-to-new-research/> 20.
- <https://www.guru3d.com/story/microsoft-reportedly-readies-billion-bid-to-acquire-valve-steam/> 21.
- <https://www.reuters.com/technology/eu-data-protection-board-says-chatgpt-still-not-meeting-data-accuracy-standards-2024-05-24/> 23.
- <https://www.wired.com/story/anthropic-black-box-ai-research-neurons-features/> 24.
- <https://www.marktechpost.com/2024/05/21/gradient-ai-introduces-llama-3-8b-gradient-instruct-1048k-setting-new-standards-in-long-context-ai/> 25.
- <https://decrypt.co/232299/mistral-ai-7b-upgrade-uncensored-cohere-aya-open-source> 26.
- <https://www.phoronix.com/news/Llamafire-0.8.5-Released> 27.
- <https://www.bing.com/chat?showconv=1> 28.
- <https://huggingface.co/spaces/ehristoforu/dalle-3-xl-lora-v2> 29.
- <https://www.marktechpost.com/2024/05/28/nv-embed-nvidias-groundbreaking-embedding-model-dominates-mteb-benchmarks/> 30.
- <https://www.marktechpost.com/2024/05/31/autocoder-the-first-large-language-model-to-surpass-gpt-4-turbo-april-2024-and-gpt-4o-in-pass1-on-the-human-eval-benchmark-test-90-9-vs-90-2/> 31.
- https://huggingface.co/Bin12345/AutoCoder_S_6.7B 32. <https://opencodeinterpreter.github.io/> 33.
- <https://www.marktechpost.com/2024/05/26/finrobot-a-novel-open-source-ai-agent-platform-supporting-multiple-financially-specialized-ai-agents-powered-by-llms/> 34.
- <https://github.com/ai4finance-foundation/finrobot> 35. <https://ai4finance.org/projects> 36.
- <https://github.com/AI4Finance-Foundation/FinGPT> 37.
- <https://www.marktechpost.com/2024/06/01/llamafs-an-open-source-self-organizing-file-system-with-llama-3/> 38.
- <https://www.marktechpost.com/2024/06/01/rag-me-up-a-generic-ai-framework-server-uis-that-enables-you-to-do-rag-on-your-own-dataset-easily/?amp> 39.
- <https://huggingface.co/mistralai/Codestral-22B-v0.1> 40. <https://huggingface.co/nvidia/NV-Embed-v1>

<https://github.com/ludouzi/fooyin?tab=readme-ov-file>

<https://civitai.com/images/7130072>

11. <https://openai.com/index/dall-e-3/> 16.
- https://www.google.com/search?q=rag+ai+github&client=ms-android-oneplus-rvo3&sca_esv=82d30ddfe6a51a7e&sxsrf=ADLYWILD6MIZAV8TErzsYRpTbmOScd725A%3A1717893096062&ei=6PdkZoC4A

[4mihbIP_tSxmQY&oq=rag+ai+github&gs_lp=EhNtb2JpbGUtZ3dzLXdpei1zZXJwlg1yYWcgYWkgZ2l0aHViMgUQABiABDIIEAAYFhgKGB4yCBAAGIAEGKIEMggQABiABBiiBDIIEAAYgAQYogQyCBAAGIAEGKIEMggQABiABBiiBEijLVCBCVidGnABeAGQAQCYAdQHoAGAF6oBCTAuNC41LTEuMrgBA8gBAPgBAZgCCKACnBjCAgoQABiwAxjWBBhHwgINEAAYgAQYsAMYQxiKBcICGRAuGIAEGLADGNEDGEMYxwEYyAMYigXYAQHCAhMQLhiABBiwAxhDGMgDGloF2AEBwglGEAAYFhgewglFECEYoAGYAwCIBgGQBg-6BgQIARglkgcLMS4zLjEuNS0xLjKgB-0S&client=mobile-gws-wiz-serp#ip=1](https://www.google.com/search?q=rag+ai+github&gs_lp=EhNtb2JpbGUtZ3dzLXdpei1zZXJwlg1yYWcgYWkgZ2l0aHViMgUQABiABDIIEAAYFhgKGB4yCBAAGIAEGKIEMggQABiABBiiBDIIEAAYgAQYogQyCBAAGIAEGKIEMggQABiABBiiBEijLVCBCVidGnABeAGQAQCYAdQHoAGAF6oBCTAuNC41LTEuMrgBA8gBAPgBAZgCCKACnBjCAgoQABiwAxjWBBhHwgINEAAYgAQYsAMYQxiKBcICGRAuGIAEGLADGNEDGEMYxwEYyAMYigXYAQHCAhMQLhiABBiwAxhDGMgDGloF2AEBwglGEAAYFhgewglFECEYoAGYAwCIBgGQBg-6BgQIARglkgcLMS4zLjEuNS0xLjKgB-0S&client=mobile-gws-wiz-serp#ip=1) 17.

<https://github.com/langchain-ai/rag-from-scratch> 18. <https://github.com/mrdbourke/rag-resources> 19. <https://neurosciencenews.com/ai-llm-reasoning-26248/> 20. <https://decrypt.co/233512/how-to-create-photorealistic-ai-pictures-with-stable-diffusion?amp=1> 21. <https://github.com/infiniflow/ragflow/tree/main?tab=readme-ov-file> 22. https://www.reddit.com/r/Diablo/comments/185lyro/diablo_4_on_steam_deck_steam/ 23. <https://overkill.wtf/how-to-install-the-diablo-iv-beta-on-steam-deck/#:~:text=Open%20up%20desktop%20Steam%20%2C%20go,net%20installer%20file%20to%20Steam.> 24. <https://venturebeat.com/> 25. <https://venturebeat.com/ai/chinas-deepseek-coder-becomes-first-open-source-coding-model-to-beat-gpt-4-turbo/> 26. <https://venturebeat.com/ai/when-to-ignore-and-believe-the-ai-hype-cycle/> 27. <https://venturebeat.com/ai/agi-isnt-here-yet-how-to-make-informed-strategic-decisions-in-the-meantime/> 28. <https://venturebeat.com/ai/new-transformer-architecture-could-enable-powerful-llms-without-gpus/> 30. <https://github.com/infiniflow/ragflow/commits/main/> 31. <https://github.com/ggozad/oterm/issues?q=is%3Aissue+is%3Aclosed> 32. https://en.m.wikipedia.org/wiki/The_Sims_Mobile 33. https://www.google.com/search?gs_ssp=ejzj4tTP1Tcws7DMtTBg9OLOtUxRKCITSE_MTQUATfoG1w&q=mad+tv+game&oq=mad+tv+game&gs_lcrp=EgZjaHJvbWUqBwgBEC4YgAQyCggAEAAAY4wIYgAQyBwgBEC4YgAQyBwgCEAAAYgAQyBwgDEAAAYgAQyBwgEEAAAYgAQyCAGFEAAAYFhgeMggIBhAAGBYHjIICAcQABgWGB4yCAGIEAAAYFhgeMggICRAAGBYHjIICAOQABgWGB4yCAGLEAAAYFhgeMggIDBAAGBYHjIICAOQABgWGB4yCAGOEAAAYFhge0gEINDQyMGowajSoAg6wAgE&client=ms-android-oneplus-rvo3&sourceid=chrome-mobile&ie=UTF-8 34. <https://en.m.wikipedia.org/wiki/SimAnt> 35. [https://en.m.wikipedia.org/wiki/Spore_\(2008_video_game\)](https://en.m.wikipedia.org/wiki/Spore_(2008_video_game)) 36. <https://en.m.wikipedia.org/wiki/Darkspore> 37. <https://en.m.wikipedia.org/wiki/SimTower> 38. <https://en.m.wikipedia.org/wiki/SimLife> 40. <https://techxplore.com/news/2024-06-technique-capabilities-large-language.html> 54. <https://discourse.beets.io/t/beetle-a-web-interface-for-beets/1678> 55. <https://gitlab.com/maxburon/beetle> 56. <https://github.com/kdelwat/beets-ui?tab=readme-ov-file> 57. <https://discourse.beets.io/t/web-gui-that-displays-album-art/2160> 58. <https://github.com/MrNuggelz/beets-web-import> 59. <https://www.visophyte.org/blog/2014/04/06/webpd-a-polymer-based-web-ui-for-the-beets-music-library-manager/> 60. <https://hub.docker.com/r/linuxserver/beets> 61. <https://pypi.org/project/beets-web-import/> 62. <https://community.roonlabs.com/t/beets-to-manage-library/190566> 63. <https://huggingface.co/stabilityai/stable-diffusion-3-medium> 64. <https://huggingface.co/spaces/stabilityai/stable-diffusion-3-medium> 65. <https://stable-diffusion-art.com/stable-diffusion-3-local/> 66. https://www.google.com/search?q=ollama+summarize+document&oq=ollama+summar&gs_lcrp=EgZjaHJvbWUqBwgAEAAAYgAQyBwgAEAAAYgAQyBggBEEUYOTIHCAIQABiABDIHCAMQABiABDIHCAQQABiABDIHCAUQABiABDIHCAYQABiABDINCAcQABiGAXiABBiKBTINCAgQABiGAXiABBiKBTINCAkQABiGAXiABBiKBTIKCAoQABiABBiiBDIKCAsQABiABBiiBNIBCDY2MTNqMGo3qAIUsAIB&client=ms-android-oneplus-rvo3&sourceid=chrome-mobile&ie=UTF-8#ip=1 67. <https://www.sheshbabu.com/posts/system-wide-text-summarization-using-ollama-and-applescript/> 68. <https://github.com/open-webui/open-webui> 69. <https://docs.openwebui.com/getting-started/env-configuration/#web-search> 70. <https://medium.com/free-or-open-source-software/ollama-llava-describe-and-summarise-websites-453>

2ef1b1d3a 71.

<https://discourse.devontechnologies.com/t/experimenting-with-llama2-llm-for-local-file-classification-renaming-summarizing-analysing/76945> 72.

<https://www.marqo.ai/courses/fine-tuning-embedding-models> 73.

<https://diptimanrc.medium.com/text2sql-opensource-duckdb-nsql-7b-with-ollama-and-llamaindex-on-local-setup-6f266f78bc4f> 75. <https://docs.openwebui.com/features> 76.

<https://github.com/open-webui/open-webui/discussions/1349> 77.

<https://github.com/open-webui/open-webui> 78.

<https://medium.com/@kelvincampelo/how-ive-optimized-document-interactions-with-open-webui-and-rag-a-comprehensive-guide-65d1221729eb> 79. <https://docs.openwebui.com/tutorial/rag/> 80.

<https://docs.openwebui.com/pipelines/> 81.

https://www.reddit.com/r/LocalLLaMA/comments/1cm6u9f/local_web_ui_with_actually_decent_rag/ 82.

<https://www.linkedin.com/pulse/fine-tuning-open-source-llms-webui-paul-hankin-rajve> 83.

<https://github.com/open-webui/pipelines> 84. <https://github.com/open-webui/pipelines/discussions/62>

85. <https://ikasten.io/2024/06/03/getting-started-with-openwebui-pipelines/> 86.

<https://github.com/open-webui/pipelines?tab=readme-ov-file#-quick-start-with-docker> 87.

<https://m.youtube.com/watch?v=NrJEFrth27Q&pp=ygUVbWFyaWFoIGNhcmV5IGVtb3Rpb25z> 88.

<https://github.com/yt-dlp/yt-dlp> 90. <https://github.com/nathom/streamrip> 91.

<https://community.spiceworks.com/t/domain-password-computer-password-do-not-match/153825>

1. <https://venturebeat.com/> 2.

<https://venturebeat.com/ai/groq-open-source-llama-ai-model-tops-leaderboard-outperforming-gpt-4o-and-claude-in-function-calling/> 3.

<https://venturebeat.com/ai/hugging-faces-smollm-models-bring-powerful-ai-to-your-phone-no-cloud-required/> 4. <https://huggingface.co/blog/smollm> 5. <https://huggingface.co/HuggingFaceTB/SmolLM-135M>

6.

<https://venturebeat.com/ai/microsofts-designer-ai-image-generator-comes-to-ios-android-with-new-features/> 7.

<https://venturebeat.com/ai/mistral-releases-codestral-mamba-for-faster-longer-code-generation/> 8.

<https://venturebeat.com/ai/microsofts-new-ai-system-spreadsheetllm-unlocks-insights-from-spreadsheets-boosting-enterprise-productivity/> 9.

<https://venturebeat.com/ai/gen-ai-takes-over-finance-the-leading-applications-and-their-challenges/>

10. <https://github.com/gpt-engineer-org/gpt-engineer?tab=readme-ov-file> 11.

<https://pyimagesearch.com/2024/07/01/exploring-oobabooga-text-generation-web-ui-installation-features-and-fine-tuning-llama-model-with-lora/> 12.

https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard 13.

https://www.reddit.com/r/LocalLLaMA/comments/18j39qt/what_embedding_models_are_you_using_for_rag/ 14. <https://huggingface.co/spaces/mteb/leaderboard> 15. <https://tiger-ai-lab.github.io/LongRAG/>

16.

<https://venturebeat.com/ai/meta-researchers-distill-system-2-thinking-into-llms-improving-performance-on-complex-reasoning/> 17. <https://github.com/open-webui/open-webui/issues> 18.

<https://www.marktechpost.com/2024/07/09/nvidia-introduces-rankrag-a-novel-rag-framework-that-instruction-tunes-a-single-llm-for-the-dual-purposes-of-top-k-context-ranking-and-answer-generation-in-rag/?amp> 19. <https://ollama.com/library/llama3-chatqa> 20.

<https://huggingface.co/nvidia/Llama3-ChatQA-1.5-70B> 21. <https://huggingface.co/nvidia/NV-Embed-v1>

22. <https://chatqa-project.github.io/> 23.

<https://www.marktechpost.com/2024/07/10/plandex-a-reliable-and-developer-friendly-ai-coding-agent-in-your-terminal/?amp> 24. <https://github.com/plandex-ai/plandex?tab=readme-ov-file> 25.

<https://docs.plandex.ai/hosting/self-hosting> 26. <https://neo4j.com/blog/graphrag-manifesto/> 27.

<https://ollama.com/library/deepseek-v2/tags> 28.

<https://www.techspot.com/news/103778-google-opening-dark-web-reports-all-free-google.html> 29.

<https://www.ontotext.com/knowledgehub/fundamentals/what-is-graph-rag/> 30.
<https://techxplore.com/news/2024-07-generative-ai-complex-tabular.html> 31.
<https://www.marktechpost.com/2024/07/07/internlm2-5-7b-chat-open-sourcing-large-language-model-s-with-unmatched-reasoning-long-context-handling-and-enhanced-tool-use/?amp> 32.
<https://arxiv.org/abs/2407.01219> 33.
<https://github.com/open-webui/open-webui/discussions/2146#discussioncomment-10073679> 34.
<https://ollama.com/library/llama3-groq-tool-use> 35.
<https://github.com/open-webui/open-webui/pull/3449> 36.
<https://github.com/open-webui/open-webui/discussions/3522>

From:

<http://wuff.dyndns.org/> - **Wulf's Various Things**

Permanent link:

<http://wuff.dyndns.org/doku.php?id=temp:bookmarks>

Last update: **2024/07/22 00:20**

